

## ABSTRACT

A new recommendation technique, referred to as "personality diagnosis", that can be seen as a hybrid between memory-based and model-based collaborative filtering techniques, is described. Using personality diagnosis, all data may be maintained throughout the processes, new data can be added incrementally, and predictions have meaningful probabilistic semantics.

Each entity's (e.g., user's) reported attributes (e.g., item ratings or preferences) may be interpreted as a manifestation of their underlying personality type. Personality type may be encoded simply as a vector of the entity's (e.g., user's) "true" values (e.g., ratings) for attributes (e.g., items) in the database. It may be assumed that entities (e.g., users) report values (e.g., ratings) with a distributed (e.g., Gaussian) error. Given an active entity's (e.g., user's) known attribute values (e.g., item ratings), the probability that they have the same personality type as every other entity (e.g., user) may be determined. Then, the probability that they will have a given value (e.g., rating) for a valueless (e.g., unrated) attribute (e.g., item) may then be determined based on the entity's (e.g., user's) personality type. The probabilistic determinations may be used to determine expected value of information. Such an expected value of information could be used in at least two ways. First, an interactive recommender could use expected value of information to favorably order queries for attribute values (e.g., item ratings),

- thereby mollifying what could otherwise be a tedious and frustrating process. Second, expected value of
- information could be used to determine which entries of a database to prune or ignore -- that is, which entries,
- 5 which if removed, would have a minimal effect of the accuracy of recommendations.